American Council on Science and Health
47 Maple Street
Summit, NJ 07901

Of Mice and Men:

The Benefits and Limitations of Animal Cancer Tests

A Report by the American Council on Science and Health

Of Mice and Men:

The Benefits and Limitations of Animal Cancer Tests

This report on animal cancer tests was written by **William R. Havender, Ph.D.**, of Berkeley, California, a consultant on environmental carcinogens and a Scientific Advisor to the **American Council on Science and Health.**

ACSH gratefully acknowledges the comments and contributions of the following individuals who reviewed this report:

Stephen Barrett, M.D.
Allentown, Pennsylvania

Joseph F. Borzelleca, Ph.D.
Medical College of Virginia

F. M. Clydesdale, Ph.D.
University of Massachusetts

Julius M. Coon, M.D., Ph.D.
Thomas Jefferson University

Owen R. Fennema, Ph.D.
University of Wisconsin

Dean C. Fletcher, Ph.D.
Washington State University

Ralph W. Fogleman, D.V.M.
Upper Black Eddy, Pennsylvania

F. J. Francis, Ph.D.
University of Massachusetts

Richard A. Greenberg, Ph.D.
American Council on Science and Health

Wayland J. Hayes, Jr., M.D., Ph.D.
Vanderbilt University

William T. Jarvis, Ph.D.
Loma Linda University

Manfred Kroger, Ph.D.
Pennsylvania State University

Howard D. Maccabee, Ph.D., M.D.
Walnut Creek, California

Roger P. Maickel, Ph.D.
Purdue University

Cover photo courtesy of Charles River Breeding Laboratories, Inc.

Kathleen A. Meister, M.S.
American Council on Science and Health

Joseph M. Miller, M.D., M.P.H.
New Hampton, New Hampshire

David B. Roll, Ph.D.
University of Utah

Paul D. Saltman, Ph.D.
University of California, San Diego

Herbert P. Sarett, Ph.D.
Sarasota, Florida

Sidney Shindell, M.D., LL.B.
Medical College of Wisconsin

Fredrick J. Stare, M.D., Ph.D.
Harvard School of Public Health

Stephen S. Sternberg, M.D.
Memorial Sloan-Kettering Cancer Center

Elizabeth M. Whelan, Sc.D., M.P.H.
American Council on Science and Health

Philip L. White, Sc.D.
American Medical Association

John A. Zapp, Jr., Ph.D.
Kennett Square, Pennsylvania

The American Council on Science and Health (ACSH) is a national consumer education association directed and advised by a panel of scientists from a variety of disciplines. **ACSH** is committed to providing consumers with scientifically balanced evaluations of issues relating to food, chemicals, the environment, and health.

ACSH is a nonprofit association exempt from federal income tax under Section 501(c)(3) of the Internal Revenue Code. All contributions are tax-deductible as provided by law.

Individual copies of this report are available at a cost of $2.00. Prices for 10 or more copies are available on request.

March 1984

Introduction

It began with cranberries. They were conspicuously absent from our Thanksgiving dinner tables in 1959 because it had been announced that an herbicide used in the cranberry bogs could give us cancer.

Then followed cyclamates, Red Dye #2, saccharin, nitrites, and others. It seems that almost every month we hear that another food ingredient has been found to be carcinogenic (cancer-causing). Or perhaps we hear that the ground water in the area where we live is contaminated with pesticides, or with dioxin or other chemical wastes, and that these substances pose a cancer threat. Or that the home insulation that we installed a few years back in response to the national call to conserve energy is leaking fumes, and that those fumes are a cancer hazard.

All of these cancer scares, and many others, have been triggered by the same type of alarm: tests in laboratory animals which had led to a pronouncement that the substance in question could cause cancer in people. These tests affect all of us. Our health depends on decisions which are based on them. The availability of products that we use every day depends on them. And the flow of billions of dollars in pollution controls, insurance premiums, product reformulations, damage payments, and legal fees depends upon them.

Is this heavy reliance on animal cancer testing justified? There are good reasons to ask this question. After all, these all-important tests are performed on laboratory animals, not on people. The conditions used in the tests, particularly the administration of huge doses of a suspect substance, bear little resemblance to realistic conditions of human exposure. Are the results of such tests truly applicable to human health?

Most people first became aware of this issue when the Food and Drug Administration (FDA) declared saccharin a carcinogen in 1977. This determination was based on tests showing that the artificial sweetener could cause cancer in rats under conditions of massive, prolonged exposure. Tumors were found only when saccharin was administered in amounts equivalent to the consumption of about a thousand cans of diet soda a day by a human, commencing with the weaning of the *parent* generation of rats, and continuing through the conception, gestation, nursing, and adult lifetime of the animals that ultimately developed tumors. In fact, even under these extreme conditions, only male rats developed tumors, only one organ was affected (the bladder), and the tumors were not observed to be metastatic (spreading to distant sites) or otherwise lethal. Many people who learned these facts understandably questioned whether this result really implied that the consumption of ordinary amounts of saccharin by humans posed a significant cancer hazard.

On the other hand, U.S. health and safety regulatory agencies constantly assure us that the results of these tests are "valid" and provide a sound basis for decisions about human hazards.

In this report, the American Council on Science and Health (ASCH) examines the controversy surrounding the use of animal cancer tests and looks at their benefits and limitations.

Why is so much animal cancer testing conducted?

Extensive carcinogenicity testing is conducted because cancer is a prevalent disease in our society and because most cancer is believed to be due to "extrinsic" factors, that is, to factors other than differences in genetic susceptibility to the disease. If extrinsic cancer-causing agents can be identified and if exposure to these agents can be reduced, it should be possible to prevent many cases of human cancer.

How prevalent is cancer? Is it on the increase?

Some 20 to 25 percent of Americans can expect to develop cancer at some point in their lives. This proportion is much higher than used to be the case, but the increase is not due to a sudden epidemic of cancer. Rather, it stems primarily from the great reduction in the number of people dying from infectious diseases such as tuberculosis, diphtheria, pneumonia, smallpox, and polio. Cancer is mainly a disease of older adults. Thanks to the control of the infectious diseases that once killed many young people, most Americans can expect to live to an age when cancer is more likely to develop.

Thus, cancer statistics, to be meaningful, must be corrected for the increase in the number of people living into old age. Such "age-adjusted" statistics show, with one exception, that the overall incidences of the major types of cancer have been rather steady or even decreasing since the 1930s (stomach cancer in particular has dropped dramatically) when nationwide statistics first became available. The exception is a large rise in respiratory cancer, particularly lung cancer. The primary cause of this increase—cigarette smoking—is well known.

Why do scientists believe that external factors play an important role in the occurence of cancer?

The rates of particular types of cancer vary greatly from place to place around the world and from time to time. Liver cancer, for example, is common in parts of sub-Saharan Africa but is rare in the United States. Breast cancer is common in the U.S. but rare in Japan, while stomach cancer is common there and uncommon here.

That differences of this type are not due to genetic differences in susceptibility can be shown by the experiences of migrant populations and by instances in which cancer rates have changed very rapidly. Descendants of Japanese immigrants to the United States develop the cancer incidence patterns typical of other Americans within one generation. Persons of Scandinavian and Celtic descent have low rates of skin cancer when living in Northern Europe, but persons of the same ethnic backgrounds who live in tropical areas have high skin cancer rates, due to greater exposure to sunlight.

Lung cancer was a very rare disease in the U.S. in 1900. It is now the leading cause of cancer death among American men and the second-leading cause among American women. Stomach cancer was a common cause of cancer death early in this century but now is rare. Genetics cannot account for such drastic changes in just a few generations, but changing exposure to extrinsic cancer-causing agents can.

It is currently estimated that between 75 and 80 percent of human cancer in the U.S. is attributable to extrinsic factors and hence is potentially preventable. However, the fact that, in general, nonrespiratory cancer rates have been steady or declining for many decades suggests that, whatever the major extrinsic causes of current nonrespiratory cancers are, they are long-established, not new, aspects of our way of life.

If the detection of agents that can cause human cancer is so important, shouldn't humans, rather than animals, be studied?

Humans *are* studied in an effort to detect factors that affect the risk of cancer. This is the province of epidemiology: the science that examines the patterns of occurrence of human disease and of exposure to suspect causative agents.

Epidemiology has the great virtue of directly identifying *human* risk factors, and hence it does not suffer from the same kinds of uncertainties of interpretation that are associated with animal tests. Many chemicals and industrial processes have been found by this means to cause human cancer. The International Agency for Research on Cancer (IARC) recently reported 18 such causes, in addition to cigarette smoking, alcohol, and radiation, which had already been identified as cancer-causing agents. The same report also listed an additional 18 "probable" causes of human cancer. Another recent survey identified some 40 risk factors for various human cancers, including lifestyle factors such as "late age at first preganancy," "sexual promiscuity," "obesity," and certain infectious diseases (such as hepatitis B) and parasitic infections (such as that caused by *Schistosoma haemotobium*) which predispose a person to develop particular types of cancer.

Epidemiology suffers from several inherent defects, however. It is difficult to establish small effects with statistical confidence by epidemiological means. "Small," in this case, means proportionately small; an effect that could not be confirmed by an epidemiologic

study of a limited group of people might still be responsible for a significant number of cases of disease each year in a larger population.

One problem in epidemiologic studies is the difficulty of assembling reliable information on large numbers of people by means of interviews or examination of medical records. Another is the difficulty of finding groups of people to compare who differ *only* in the single factor under study. If they differ in other ways, these other differences, called "confounding variables," might generate a spurious relationship or conceal a true one. This is often a problem even when the effect under investigation is large, but when it is at most small, such as the hypothetical effect of saccharin on bladder cancer or of hair dyes on breast cancer, the problem can be insurmountable.

It is particularly difficult to find comparison groups of people who have never been exposed to the factor under study (zero-dose control groups). For example, if we wanted to test the hypothesis that caffeine caused cancer, it would be almost impossible to find a sizable group of people with no exposure to caffeine at all, since caffeine is a constituent of coffee, tea, cola beverages and some other drinks; cocoa and chocolate products; and many over-the-counter drugs. We would have to look at special groups of people such as Seventh Day Adventists, who make a deliberate effort to abstain from major sources of caffeine (such as coffee). However, groups of this type differ from the general population in many other ways, so we would then be faced with many possible confounding variables.

Epidemiological studies led to the very important discovery that cancer can first appear decades *after* initial exposure to a carcinogenic substance. For instance, cigarette smokers don't usually develop lung cancer sooner then twenty years after they start smoking. Lung cancers in asbestos workers also typically develop several decades after the first exposure to asbestos. Some of the daughters of women who were given the drug diethylstilbestrol (DES) during pregnancy have developed vaginal cancer. These cancers only became manifest after the daughters passed puberty, many years following the actual exposure to the drug.

Such long latency periods are another factor that makes epidemiological investigations of cancer causation difficult, because the search for causes of current cancer cases must focus on events that took place several decades in the past. Memories of these events have faded and records have often been lost.

Moreover, these long latencies mean that one must wait for decades to establish that a current or recent exposure will or will not result in cancer in the future.

Thus, human studies cannot assess the effects of the many new chemicals constantly generated by a modern industrial economy—about a thousand each year in the U.S. alone—until sufficient time has elapsed. Since the toxic effects, including cancer, of such new chemicals must be identified before significant human exposure to them can be allowed, altenative testing methods must be used.

What is the best alternative when epidemiology is not appropriate?

Animal cancer testing, where laboratory animals serve as proxies for humans, is the best alternative we have. Rats and mice are usually selected in preference to other species such as monkeys that might more closely resemble humans because rodents are small and comparatively inexpensive to maintain, and because they are short-lived, allowing lifetime studies to be done on them in a reasonable time.

It was first demonstrated in 1915 that cancer could be deliberately induced in animals by treating them with a chemical. A wide variety of methods of animal cancer testing have been used since that time. Chemicals have been introduced into experimental animals by every orifice (orally, nasally, urethrally, vaginally, rectally), by various types of injection (intramuscular, intraperitoneal, intravenous, subcutaneous), by skin painting, by surgery, and by other methods. Many variations in experimental design, with different dosages, lengths and patterns of dosing, and different observation periods have been used. The thoroughness of the pathological examinations for tumors has also varied tremendously.

In the 1960s, an experimental design was developed that was believed to be best suited for the routine, rigorous screening of large numbers of chemicals for carcinogenicity. This design was adopted by the National Cancer Institute (NCI) for its large carcinogen screening program (NCI Bioassay), and it is currently used by the National Toxicology Program (NTP), which has taken over this program from NCI.

How are these standard animal cancer tests done?

A typical test is performed on both sexes of two species of animals, chiefly rats and mice. The animals are exposed to the chemical for most or all of their lives (about two years) to maximize the chances of detecting cancers with long latency periods. (Latency periods in animals with short lifespans, such as rats and mice, are much shorter than they are in humans, but they still may comprise a substantial portion of the animal's lifetime.) The exposure route is selected to mimic as

closely as possible the route of human exposure. Animals are usually exposed by including the test substance in their diets, but in some cases inhalation or exposure through drinking water is more suitable. Unpalatable substances are given by stomach tube.

Normally, three dose groups of animals are used. One, the control group, does not receive the chemical but is otherwise treated indentically with the other dose groups. The other two groups receive, respectively, a "high" or a "low" dose. Fifty animals make up each sex/species/dose subgroup for a total of 600 animals.

Before the test is performed, it is necessary to do preliminary experimentation to determine the proper dose levels. At the end of the test, pathologists must examine thousands of microscope slides of some 40 of the animals' tissues and organs to detect even minute tumors that might not be seen with the naked eye. When the time necessary to perform these steps is added to the time required for the actual test, the whole procedure takes three or more years to complete. The cost is about $1000 per animal, or $600,000 for a typical test on a single chemical.

With only 600 animals, a chemical that is a relatively weak carcinogen, say, one that induces only one tumor in a thousand animals, probably would not be identified as cancer-causing. A substance would have to induce cancer in about 7 to 10 percent of the exposed animals in order for there to be a good chance that its carcinogenic action would be detected with statistical confidence in a test of this size. If spontaneous tumors appeared in the control animals as well (as commonly occurs), the test would be even less sensitive.

In theory, one could design a test involving much larger numbers of animals which would be capable of detecting weak carcinogens, but increasing the number of animals increases the cost, and very large animal experiments are logistically difficult to carry out. For example, it is hard to obtain thousands of animals of the same age and strain, and the problems of keeping track of them at each stage in the experiment can be formidable.

Tests using such very large numbers of animals are not practical for routine screening and are carried out only when there is a specific, urgent research question to be pursued. For the purpose of testing sizable numbers of chemicals simply to see whether they might be carcinogenic, the NCI/NTP type of test is about as large as is practical.

What types of dosages are used in these tests?

Large dosages are used in order not to miss even a weak carcinogen, since one capable of causing just a single

case of cancer in 1000 exposed people could theoretically cause 200,000 cases of cancer in the U.S. *if* everyone were exposed. The only way to compensate, however imperfectly, for the limited number of animals that can be used in a routine screening test is by using large doses—much, much larger, in virtually all cases, than the doses that humans would be subjected to. The rationale for this is that, in general, the incidence of tumors will increase as the dose of a carcinogen gets larger.

The logical limit to this reasoning is to include the largest dose that the animals can tolerate without dying prematurely from simply being poisoned by the substance. This dose is called the Maximum Tolerated Dose (MTD), and it is standard practice to use the MTD as the higher of the two test doses in a typical animal cancer test. The "low" dose is usually set at one half or one fourth of the MTD—still a high dose compared to most human exposures.

Belief that the use of the MTD will maximize the power of these tests to detect weak carcinogens is the principal reason why regulatory agencies and some scientists defend this practice as "valid."

But are results obtained from such tests *really* valid?

This is a matter of controversy, not only among the general public but also among the scientists who perform and interpret these tests. For despite the fact that the use of maximal doses minimizes the chance that weak carcinogens will be missed, another consideration becomes salient, namely, whether the use of the MTD, which is just slightly below the dose that would kill the animals from poisoning, can by itself predispose them to develop cancer. Perhaps the animals' resistance to disease is somehow weakened. Or perhaps the animals' ability to detoxify the test compound is simply swamped by the high dosage. Many scientists are convinced that unusual metabolic events affecting the incidence of tumors may occur in heavily stressed animals. These events may not occur at all, or at least not proportionately, at the much lower dosages typical of human exposure.

These are complex issues which have no simple solution that would be generally true for all chemicals. Instead, each chemical needs separate testing to see whether the carcinogenic response at lower doses is simply a proportionately weaker version of the response seen at the MTD, or whether something unusual happens at the latter high and nearly poisonous dose.

The only way to obtain this information is to do a full "dose response" study for the chemical in question. In a

dose response study, the usual animal cancer test is performed, but at a large number of doses rather than just one or two, and the doses range down to a level characteristic of human exposure. As mentioned, however, such a test requires large numbers of animals and is prohibitively expensive to use for routine screening—which is why exaggerated doses are used in the first place. Only a few such tests have been carried out, and these are discussed below.

Faced with this dilemma, regulatory agencies have arrived at decisions based on their desire to be prudent. Since some decision must be made on the regulation of a substance even when definitive evidence of the real risk to humans at expected exposure levels is lacking, the agencies choose to make decisions which would err, if they err at all, in the direction of public safety. For this reason, high doses are simply *assumed* not to lead to aberrant results. Whatever outcome is seen at the MTD is conjectured to exist in proportion at all lower doses as well. In other words, the assumption is made that the dose response curve is *linear* (see the top graph in Figure 1) passing through the origin (i.e., at zero dose, there would be no induced tumors). If one makes this assumption, then it is easy to make an estimate of the risk at low doses from the level of risk seen at high doses simply by reading off a diagram such as the top graph in Figure 1. This method of risk prediction is considered to be "conservative," allowing a generous margin of safety in protecting public health.

Are dose response curves really linear?

Not necessarily, and in fact not even usually. Figure 1 shows idealized versions of two possible types of dose response relationship: the linear type assumed to exist in regulators' extrapolations (top graph), and a second, curving type where the risk accelerates (i.e., gets larger at a faster and faster *rate*) as the dose gets larger (bottom graph, Figure 1). This is sometimes referred to as a "hockey stick" curve (it is actually the lower portion of what toxicologists recognize as a "sigmoid" curve). One variant of this, called a "threshold," would show an apparent "safe" dose, i.e., a dose below which no risk was evident.

Figure 2 (page 12) shows, again in idealized form, what can happen if one *assumes* that a dose response curve is linear, when in fact it has a sigmoidal shape. If the ability of this chemical to produce tumors were measured only at the dose labeled C and the dose response were assumed to be linear, one would draw a line from this point to the origin like the broken line on the graph. But if the real dose response curve were sigmoidal, like the solid line on the graph in Figure 2, then linear extrapolation would clearly *overestimate*

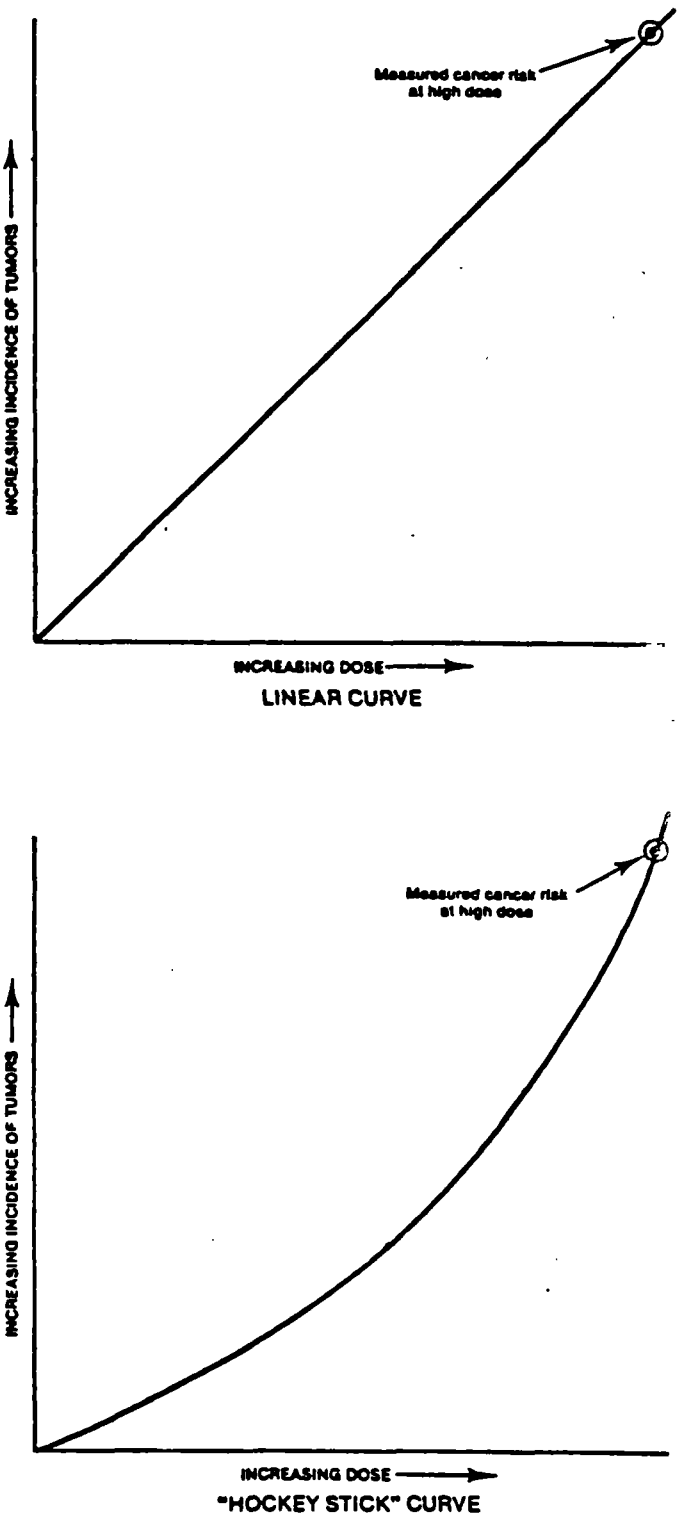

Figure 1: Two Types of Dose Response Curve

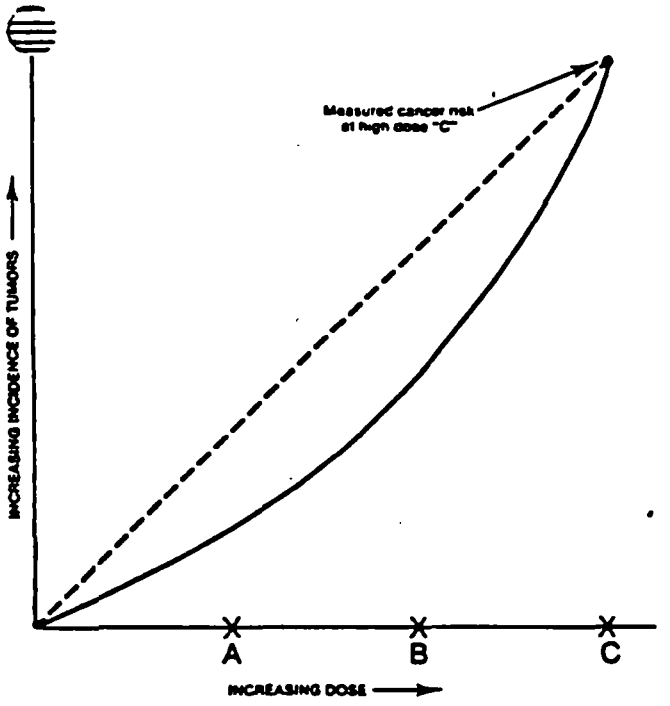

Figure 2: Dose Response Curves: Consequences of Linear Extrapolation

the actual number of tumors that would result from exposure to lower doses, such as doses A and B on the graph. This, of course, is precisely why assuming the dose response to be linear is conservative; if wrong, it is virtually certain to overstate the true risk. The trouble, however, is that such overstatements of the true low dose risk can be vast, given the huge extrapolations necessary (often more than 100-fold) in going from the MTD to the doses characteristic of human exposure.

As mentioned, some well-designed dose response tests have been carried out, and very few of them support the assumption that cancer risk is linearly proportional to dose. The largest of these tests is the famous "ED01" or "mega-mouse" study, which involved 24,000 mice and was carried out at the government's National Center for Toxicological Research. The chemical used in the test was a potent carcinogen: 2-acetylamino-fluorene. This study not only used a huge number of animals, it exposed many more animals to the low doses than the high doses, so that the small incidence of cancer expected at the low doses could be detected with statistical confidence.

Tumors were found in two different organs in the test animals: the liver and the bladder. In neither case was the dose response relationship linear. Instead, it was clearly apparent that the dose response curve for bladder cancer was the hockey stick type rather than linear, and careful statistical analysis showed the dose response for liver cancer to be nonlinear as well (although less so than for the bladder tumors).

In the ED01 study, linear extrapolation of the data for bladder cancer risk in animals exposed for two years to the highest dose would have overestimated the actual measured risk at the lowest dose (one fifth of the highest dose) by more than tenfold. Detailed examination of the liver cancer data showed that linear extrapolation of these data would also overestimate the risk at very low doses, although the error would not be as large as that with bladder tumors.

A thorough examination of all of the large, complete dose response studies that had been completed by 1981 disclosed that all but four among the 31 tests exhibited dose response curves of the sigmoidal type. More than one-third of them differed from the linear model so sharply that if the dose were decreased by a factor of five, the tumor risk would decrease by a factor of twenty-five or more, rather than by five, as would be expected under the assumption of linearity.

Formaldehyde and saccharin provide two more recent practical examples of the problems encountered with nonlinear dose response curves.

An administered dose level of formaldehyde of 14.5 parts per million (ppm) in inhaled air was shown to produce nasal cancer in 103 of 232 exposed rats. However, a level of 5.6 ppm led to only 2 nasal cancers in 235 exposed rats. Here a decrease in dosage of only threefold led to a roughly fiftyfold decrease in cancer incidence! Linear extrapolation from the risk seen at the high dose over only a *threefold* decrease in dose would lead in this instance to a fifteenfold overestimate of the true risk. Consideration of the strongly nonlinear nature of this dose response curve is thus important in assessing the true risk to humans from small exposures to formaldehyde, e.g., from the small amounts of formaldehyde that may leak from urea-formaldehyde foam insulation. Yet, in 1982, the Consumer Product Safety Commission (CPSC) banned this insulation on the basis of linear extrapolation of the high dose nasal cancer data. This ban was later overturned in the courts.

Three tests on saccharin using an unusual "two generation" design resulted in bladder tumors in male rats of the Sprague-Dawley strain. A fourth test involving the

ne sensitive species, strain, sex, and two generation esign was recently completed, but unlike the others, it was planned specifically as a dose response study. Many doses were administered, and, as with the ED01 study, much larger numbers of rats received the small doses than the large ones. As in the previous studies, tumors of the bladder were produced at the higher doses. But the shape of the dose response curve was not linear; it clearly was an accelerating, sigmoidal curve. The "best fit" curve through the various data points generated by the experiment indicated that a 100-fold decrease in dose would result in a *one million fold* decrease in tumor risk, rather than the 100-fold decrease that would be expected if the dose response relationship were linear. This shows how large the overstatement of low dose risk can be when extrapolations under the assumption of linearity must be made over very large dose ranges.

Results such as these, which are turning up more and more often, cast suspicion on the reliability of such extrapolations, which are the norm in comparing screening tests done at the MTD with typical human exposures.

How do scientists account for nonlinear dose response curves?

To understand this, it is necessary to know a little about how chemicals can act as cancer-causing agents in the body.

Most of the chemicals that induce cancer appear to act by damaging the genetic material (DNA) of cells. In theory, this might be accomplished by a single molecule of a DNA-damaging chemical, but of course it first has to get inside cells *to* the DNA, which is sequestered in the nuclei of all living cells. To do this, a chemical has to be absorbed by the body through the lungs, the digestive tract, or other means, just like oxygen, water, and nutrients must be. And like them, DNA-damaging chemicals are subject to the body's uptake, transport, biotransformation (activation or detoxification), distribution, and excretion mechanisms. Indeed, many chemicals are not carcinogenic in themselves but must be changed by the body into "activated" forms before they are able to damage DNA. These various pharmacological processes can be nonlinear with respect to administered dose.

In particular, the body's normal means of handling a specific substance can be overloaded at high doses, and the overflow may be handled in other, quite different ways. Thus, the array of metabolites of an administered chemical seen at high dose may well be different in nature and/or relative amounts from that which predominates at normal physiological levels. If this array of high dose metabolites happens to include activated, DNA-damaging forms of the chemical, or if the body is unable to excrete activated carcinogens as efficiently at maximal doses as at normal levels, then the common experimental finding that the carcinogenic effort of chemicals is disproportionately more powerful at high doses than at low doses could be explained.

There are other sorts of carcinogens, however, that do not seem to act by means of damaging DNA. These are called "nongenotoxic" or "epigenetic" carcinogens. One class of these is called "promoters," because they do not seem to be fully potent carcinogens in themselves but do have the ability to enhance, or to "promote," the carcinogenic activity of others. Some promoters may act by facilitating the entry of DNA-damaging chemicals into the cell. Others appear to act at later stages in the development of cancer after the initial DNA damage has occurred, such as by irritating tissues and causing their component cells to multiply faster. Such extra multiplication seems to increase the chances for the carcinogenic expression of already damaged DNA. Some nongenotoxic carcinogens, such as DES, act hormonally. For all of these nongenotoxic carcinogens, there is no reason even in theory for believing that single molecules might be able to cause cancer by means of these indirect mechanisms. These effects seem more akin to ordinary toxicity, and hence it is probably seriously invalid to attempt to estimate the low dose risk by means of a linear extrapolation of the carcinogenic risk visible only at high dose.

Other chemicals showing carcinogenic activity at high doses are *normal* components of the body's biochemistry (such as formaldehyde) and for this reason seem unlikely to present a carcinogenic hazard at doses comparable to that which the body itself generates every day. And others, like vitamins A and D, are normal, indeed, *necessary*, constituents of our diet: again it seems likely that our bodies can handle physiologically normal amounts of these substances without hazard.

In light of this information, why do government regulatory agencies still use linear extrapolation?

For two reasons. First, much of the information discussed above is relatively new, and it takes a while for regulatory policies to adjust to new findings.

And second, even though a number of dose response studies have yielded curves with pronounced nonlinearity, such tests have been carried out on only a relatively small number of chemicals. Enough have been done now to make plausible the hypothesis that dose response relationships are, in general, nonlinear, but not enough to prove it.

true, we can't be sure that nonlinear dose response curves will always be seen for every chemical, and in such a circumstance, the regulatory agencies have felt it prudent to adhere to the old policy of assuming a linear dose response. Of course, good experimental data about a chemical's actual dose response, when available, should displace an assumed linearity in wise policy making. Every effort should be made to obtain such information before important regulations are finalized, and to revise older regulations as new evidence becomes available.

If all or most dose response curves for carcinogens prove to be nonlinear, how might this alter the setting of regulatory policy?

If this turns out to be true, it is possible that the methods now used to protect the public from the adverse health effects of chemicals other than carcinogens will be applicable to carcinogens as well.

The general procedure for *noncarcinogenic* toxic substances has been to perform a dose response study in animals, establish the largest dose at which no observable adverse health effect occurs, and then apply a large safety factor when determining the acceptable exposure level for people. Typically, a safety factor of 100 is used for noncarcinogenic chemicals in food. This means that the acceptable human exposure level would be one-hundredth of the highest known level at which no observable adverse effect occurs in animals.

A large safety factor is used to allow for possible metabolic differences between the test animal species and humans, and also for the much greater genetic and physiological variability in human populations (old people, young adults, children, fetuses, the ill, the over- or under-nourished, etc.) compared with laboratory animals (which are typically genetically uniform, of the same age, otherwise healthy, and maintained under constant environmental conditions).

This is how "acceptable" dietary levels (Acceptable Daily Intakes, or ADI's) have been set for decades for toxic chemicals other than carcinogens, and the method, however pragmatic and lacking in theoretical elegance, has an admirable record of success.

If ADI's for carcinogens were set in this way, an even larger safety factor would be desirable, however, because the consequences of erring in determining the ADI are greater. If an ADI for a noncarcinogenic toxic substance were set too high, then sick people would start turning up rather quickly, and we could promptly correct it. In contrast, because cancer has a long latency period, any mistake made in setting an ADI for a carcinogen would probably not become apparent for decades, during which time large numbers of people would be exposed.

Why hasn't such a method been used in the past to determine acceptable levels for carcinogens?

A similar method, involving the setting of tolerance levels on the basis of the degree of risk seen in animal tests, is used in many areas of regulation. The Occupational Safety and Health Administration (OSHA), Environmental Protection Agency (EPA) and Consumer Product Safety Commission (CPSC) all set levels of allowable exposure to carcinogens. These levels are believed to present negligible risks to workers, consumers, and the general public. In addition, the FDA uses this approach for carcinogens occurring *naturally* in foods. For instance, certain foods can become contaminated with a potent carcinogen called aflatoxin, which is produced by a mold, *Aspergillus flavus*. The FDA sets acceptable levels of aflatoxin in these foods, which include peanuts, corn, wheat, rice, and milk (cows sometimes eat mold-infected corn).

Substances *deliberately added* to foods, however, are governed by different legislation. A provision of the Federal Food, Drug and Cosmetic Act, called the Delaney Clause, specifically prohibits the deliberate addition of *any* amount of a food additive (or of two other classes of regulated substances) to food if that substance has been shown to cause cancer when ingested by animals, regardless of other considerations such as dose response relationships or the possibility that the substance might afford a health benefit in its normal uses (e.g., saccharin is believed by many to help in controlling weight and managing diabetes.

Does the Delaney Clause have any scientific basis?

Some scientists and regulatory officials have believed on theoretical grounds that carcinogens are fundamentally different from other toxic chemicals, because the change in DNA leading to cancer produces a self-replicating altered cell, so that even one such change, produced by a single molecule of a carcinogen, could potentially result in cancer. This is in constrast to situations of ordinary toxicity, where many cells have to be injured before symptoms of illness become evident. This belief in the uniqueness of carcinogens is the primary rationale for the Delaney Clause.

Some scientists still hold to the view that any administered dose of a carcinogen, however small, poses a significant risk. But this view is coming more and more into dispute because of the increasing evidence that carcinogens often, and perhaps usually, have nonlinear dose response curves and growing awareness of the

complexity of the steps that must take place in the body before an active form of a carcinogenic chemical can reach and damage cell DNA.

What is the actual record of animal cancer tests? How well do these tests on rats and mice predict the ability of substances to cause cancer in people?

The usual answer to this question is to reverse it and to ask how many human carcinogens have been shown to be animal carcinogens. This is in fact a very different question, as we shall shortly see, but the answer to it is that all but one of the chemicals that have been shown to cause cancer in humans have also been shown to cause cancer in animals. The exception is arsenic. Thus, what we may call the "false negative" rate—the fraction of chemicals that are known human carcinogens that come out negative in well designed and well executed animal cancer tests—is very low, in the neighborhood of a few percent.

The inference is then frequently made that if the correlation in this direction is so good, then the correlation in the reverse direction must be pretty good, too—i.e., that practically all chemicals shown to be carcinogens in animals must be able to cause cancer in humans as well. This may not be true, however.

There are, to be sure, plenty of chemicals that have been judged to be carcinogens in at least one animal test for which it has not been possible to securely establish that they also cause human cancer. DDT, saccharin, and cyclamate are examples. However, this cannot provide a definitive answer to the question because, just as with animal cancer tests, epidemiological studies are limited by their size and design, and, as mentioned, their sensitivity is not high. They are only capable of detecting cancer incidences larger than the statistical limit inherent in their design. Failing to find a positive result will set a "cap," or upper boundary, on the possible incidence of cancer, but it can never establish that the incidence is actually zero. Thus, it is simply not possible to show by means of epidemiology that a chemical is *completely* incapable of causing cancer in all individuals and under all conditions. Hence, we have no way to discover what fraction of positive animal cancer test results is actually "falsely" positive vis-a-vis humans.

Are there any cases in which animal cancer tests have successfully predicted carcinogenicity in humans?

Certainly. There have been seven cases so far in which chemicals were *first* found to be carcinogenic in animal tests and were later discovered to be carcinogenic in humans. The substances are aflatoxin, 4-aminobiphenyl, DES, bis(chloromethyl)ether, melphalan, mustard gas, and vinyl chloride. The problem, though, is that this is hindsight and is selective; at the same time that these studies were done, there were also multitudes of other chemicals being judged positive in animal tests that have not as yet been confirmed as human carcinogens. Currently, several hundred chemicals have been judged to be carcinogens on the basis of animal tests, and one cannot pick just the few of these that have been vindicated as human carcinogens to argue that animal tests in general are excellent predictors of cancer risk in man. The seven successes might well have been due to chance rather than to the predictive power of animal tests. Thus, insufficient evidence has been collected to argue persuasively that animal cancer tests, as currently conducted and evaluated, can be used with confidence to predict whether a given chemical will cause cancer in humans.

This is unfortunate because regulatory agencies are left with no choice but to resolve all uncertainties in the interpretation of animal cancer tests by assuming the worst in each contested instance. Such decisions include: use of the MTD so as not to miss weak carcinogens; the assumption that risk is linearly related to dose; using the results from the most susceptible species, strain, and sex as the basis for inferring human risk; ignoring the experience of decades of safe use by humans, or other negative evidence; counting benign tumors as though they were as significant as malignant ones (on the theory that some benign tumors can progress to malignancy, and one cannot determine which, or how many, will do so); and many similar decisions made in an effort to be as prudent as possible.

The compounded effect of all of these choices is to bias the analysis of animal tests strongly in the direction of concluding that the substance in question is a carcinogen. And since, as discussed, we have no way of discovering when a prediction of human cancer risk based on the positive judgment of an animal test is false, we cannot measure the effect that this deliberate bias is having on the frequency of false positive judgments. It could be, as the regulatory advocates claim, that the net effect of all of these decisions is simply to detect more and more weak carcinogens, and that truly innocuous chemicals are only rarely, if at all, mistakenly claimed to be carcinogenic to humans. But it is also possible that this bias is raising the false positive rate to unacceptably high levels. In this case, the identification of truly dangerous substances would be swamped by false positive judgments, and the ability of regulatory agencies to make *discriminating* policy deci-

sid—i.e., identifying and controlling the risks that *matter*—would be severely hindered. It is, at present, not possible to decide between these alternatives.

If the design and interpretation of animal cancer tests is biased in favor of positive judgments, wouldn't an unexpectedly large percentage of chemicals tested in this way yield positive results?

Yes. And this is just what we see. Of 190 chemicals that had been tested and reported by the NCI as of 1980, 98 or 52 percent were judged to be carcinogens in at least one of the two species tested. NTP, in an update of these data, reported that 42 percent of 252 tested chemicals (including the 190 mentioned above) in the combined NCI/NTP series were positive in at least one species. This high proportion is quite unexpected, because the basis of the effort to identify and banish human carcinogens was that these substances were relatively uncommon; after all, if a substantial fraction of *all* of the thousands of chemicals one comes into contact with each day were carcinogens, this approach to cancer prevention would be hopeless.

Of course, the high frequency of positive results in the NCI/NTP series could be explained away by saying that these chemicals were preselected as likely carcinogens in the first place, and that NCI/NTP was so successful in their preliminary selection that about half of their candidates turned out in fact to be carcinogens. This is what NCI/NTP claims, and it may be correct. However, it could just as well be that the conditions of conducting animal tests (particularly the high doses used), combined with the prudent decision criteria used for interpreting the results of these tests, yielded a high proportion of false positive judgments.

Another unsettling finding, alluded to earlier, is that a number of normal constituents of the human diet (such as table sugar, vitamins A and D, pepper, and a mixture of egg yolks and milk) have been shown to be carcinogenic in at least one animal test. It is not impossible that these animal test results *are* validly telling us that these common foods and food constituents are carcinogens (presumably weak ones) in humans. But if so, this is indeed a very strange finding. An alternative explanation may be that these are false positives resulting from the bias in the conditions of testing and evaluation.

How well do the results of cancer tests in one animal species predict carcinogenicity in other animal species?

This is an interesting question, because if such predictions were pretty good across several species, then one could reasonably argue that the substance in question is

affecting a basic aspect of metabolism that is common to most mammals, and hence, it is likely that it poses a hazard to humans as well. Some carcinogens do, in fact, show such behavior. For instance, aflatoxin is a carcinogen in mice, rats, fish, ducks, turkeys, marmosets, tree shrews, and monkeys; 4-aminobiphenyl is a carcinogen in mice, rats, rabbits, dogs, and monkeys; asbestos is carcinogenic to mice, rats, hamsters, and rabbits; DES is carcinogenic in mice, rats, hamsters, frogs, and squirrel monkeys; 2-naphthylamine is carcinogenic to mice, hamsters, dogs, and monkeys; and benzidine is carcinogenic in rats, hamsters, and dogs. It comes as no great surprise, then, that all of these substances are also carcinogenic to humans. Indeed, of all of the chemicals identified by the IARC as carcinogenic to humans, all but a few of them have been shown to be carcinogenic in more than one animal species.
If all chemicals judged to be carcinogens on the basis of at least one animal test showed such good inter-species predictability, then one would have good reason to suppose that animal tests in general were excellent predictors of human carcinogenicity as well. However, this is not the case. Of the 190 chemicals tested on rats and mice by NCI, 98 were positive in at least one of the test species (as mentioned earlier). But of these 98, 54 were positive in *only* one species, even though comparable experimental conditions and evaluation criteria were used for both species, and the two species are closely related in an evolutionary sense. Thus about half of the chemicals were falsely positive vis-a-vis the other species. This high rate of discordance casts doubt on the validity of inferences made about human risk from rodent test results, for if predictions of carcinogenicity from one rodent species to another are wrong about half the time even when the tests are conducted under parallel conditions, there is no basis for believing that predictions made to a far more distantly related species, such as *Homo sapiens*, would be any better (or, for that matter, as good).

Considering all the difficulties involved in interpreting their results, are animal cancer tests worthless in predicting human risk?

By no means. Animal tests, when interpreted with intelligence and discrimination, can be of enormous value for assessing likely human risk. The real problem is that a policy of extreme prudence at every step in the design and interpretation of these tests obscures the very large differences in *degree* of risk that animal tests are capable of demonstrating. For example, some chemicals, like aflatoxin, are carcinogenic in virtually every species tested, and produce a high frequency of

lethal tumors in young, healthy animals at extremely minute doses. Saccharin, on the other hand, has only been shown to cause tumors in one species of the several tested, only in one sex, only under an unusual test design (the "two generation" test), only at doses some millionfold higher than the dose of aflatoxin that would be needed to induce a comparable incidence of tumors; and at that, the tumors produced were only detectable under the microscope after the animals had died of other causes. Clearly, there is a world of difference in the degree of hazard presented by these two substances, and wise public policy should recognize this difference.

The wide range in potential hazard means that in some circumstances animal test data alone *can* constitute a sound basis for severely restricting or banning the use of a chemical. If a substance causes cancer in *two or more animal species*, if it causes *highly lethal* tumors or types of tumors that *do not occur spontaneously in that kind of animal*, if the tumors appear after a *short lag time*, or if the substance causes cancer *at doses similar to* or lower than the expected levels of human exposure, then it should be viewed with great concern; and human exposure should be avoided or reduced to the lowest practicable level.

Based on such criteria, aflatoxin should certainly be regulated much more strictly than saccharin, but in fact just the opposite is the case. This is because current food safety policy regulates substances not on the degree of risk they pose but on the grounds of how and why they show up in food. Because aflatoxin is a *naturally occurring* substance in food, the FDA is empowered to set acceptable tolerance levels for it. But because saccharin is a *food additive*, the Delaney Clause governs it, and this requires that none of it be permitted in the nation's food supply. Only a series of special Acts of Congress in response to overwhelming public demand has, for the time being, delayed the implementation of a saccharin ban.

Conclusion

To sum up, the key problem does not lie so much with animal tests as with the legal and regulatory use that is made of them. The policy of exaggerated prudence has veiled the wide range in the *degree* of cancer risk that animal tests can show, leading to a lumping of major hazards to human health with minute, hypothetical ones. This has resulted in an overburdening of the regulatory process, and the consequent inability to set sensible priorities and to fashion wise policies to protect the nation's health.

American Council on Science and Health

1995 Broadway
(near 68th Street)
New York, NY 10023
Telephone: 212 362 7044

1119 19th Street, N.W.
Suite 301
Washington, DC 20036
Telephone: 202 659 8978

47 Maple Street
Summit, NJ 07901
Telephone: 201 277 0024